RAW SEQUENCE LISTING

The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.

Application Serial Number: 10/804,785
Source: EFW
Date Processed by STIC: 1-18-04

ENTERED

IFWO

RAW SEQUENCE LISTING DATE: 11/18/2004
PATENT APPLICATION: **US/10/804,785** TIME: 09:30:55

Input Set : **A:\10804785.txt**
Output Set: **N:\CRF4\11182004\J804785.raw**

<110> APPLICANT: Goedegebuur, Frits
Gualfetti, Peter
Mitchinson, Colin
Neefe, Paulien
<120> TITLE OF INVENTION: Novel CBH1 Homologs and Variant CBH1 Cellulases
<130> FILE REFERENCE: GC793-3
<140> CURRENT APPLICATION NUMBER: US 10/804,785
<141> CURRENT FILING DATE: 2004-03-19
<150> PRIOR APPLICATION NUMBER: US 60/456,368
<151> PRIOR FILING DATE: 2003-03-21
<150> PRIOR APPLICATION NUMBER: US 60/458,696
<151> PRIOR FILING DATE: 3003-03-27
<160> NUMBER OF SEQ ID NOS: 18
<170> SOFTWARE: FastSEQ for Windows Version 4.0
<210> SEQ ID NO: 1
<211> LENGTH: 1491
<212> TYPE: DNA
<213> ORGANISM: Hyprocrea jecorina
<400> SEQUENCE: 1

(ps.6)

cagtcggcct gcactctcca atcggagact cacccgcctc tgacatggca gaaatgctcg 60
tctggtggca cttgcactca acagacaggc tccgtggtca tcgacgccaa ctggcgctgg 120
actcacgcta cgaacagcag cacgaactgc tacgatggca acacttggag ctcgacccta 180
tgtcctgaca acgagacctg cgcgaagaac tgctgtctgg acggtgccgc ctacgcgtcc 240
acgtacggag ttaccacgag cggtaacagc ctctccattg gctttgtcac ccagtctgcg 300
cagaagaacg ttggcgctcg cctttacctt atggcgagcg acacgaccta ccaggaattc 360
accctgcttg gcaacgagtt ctctttcgat gttgatgttt cgcagctgcc gtgcggcttg 420
aacggagctc tctacttcgt gtccatggac gcggatggtg gcgtgagcaa gtatcccacc 480
aacaccgctg gcgccaagta cggcacgggg tactgtgaca gccagtgtcc ccgcgatctg 540
aagttcatca atggccaggc caacgttgag ggctgggagc cgtcatccaa caacgcgaac 600
acgggcattg gaggacacgg aagctgctgc tctgagatgg atatctggga ggccaactcc 660
atctccgagg ctcttacccc ccaccttgc acgactgtcg gccaggagat ctgcgagggt 720
gatgggtgcg gcggaactta ctccgataac agatatggcg gcacttgcga tcccgatggc 780
tgcgactgga acccataccg cctgggcaac accagcttct acggccctgg ctcaagcttt 840
accctcgata ccaccaagaa attgaccgtt gtcacccagt tcgagacgtc gggtgccatc 900
aaccgatact atgtccagaa tggcgtcact ttccagcagc ccaacgccga gcttggtagt 960
tactctggca acgagctcaa cgatgattac tgcacagctg aggaggcaga attcggcgga 1020
tcctctttct cagacaaggg cggcctgact cagttcaaga aggctacctc tggcggcatg 1080
gttctggtca tgagtctgtg ggatgattac tacgccaaca tgctgtggct ggactccacc 1140
tacccgacaa acgagacctc ctccacaccc ggtgccgtgc gcggaagctg ctccaccagc 1200
tccggtgtcc ctgctcaggt cgaatctcag tctcccaacg ccaaggtcac cttctccaac 1260
atcaagttcg gacccattgg cagcaccggc aaccctagcg gcggcaaccc tcccggcgga 1320
aacccgcctg gcaccaccac cacccgccgc ccagccacta ccactggaag ctctcccgga 1380

RAW SEQUENCE LISTING DATE: 11/18/2004
PATENT APPLICATION: US/10/804,785 TIME: 09:30:55

Input Set : A:\10804785.txt
Output Set: N:\CRF4\11182004\J804785.raw

```
56 cctacccagt ctcactacgg ccagtgcggc ggtattggct acagcggccc cacggtctgc    1440
57 gccagcggca caacttgcca ggtcctgaac ccttactact ctcagtgcct g             1491
59 <210> SEQ ID NO: 2
60 <211> LENGTH: 497
61 <212> TYPE: PRT
62 <213> ORGANISM: Hyprocrea jecorina
64 <400> SEQUENCE: 2
65 Gln Ser Ala Cys Thr Leu Gln Ser Glu Thr His Pro Pro Leu Thr Trp
66  1               5                  10                  15
67 Gln Lys Cys Ser Ser Gly Gly Thr Cys Thr Gln Gln Thr Gly Ser Val
68             20                  25                  30
69 Val Ile Asp Ala Asn Trp Arg Trp Thr His Ala Thr Asn Ser Ser Thr
70         35                  40                  45
71 Asn Cys Tyr Asp Gly Asn Thr Trp Ser Ser Thr Leu Cys Pro Asp Asn
72     50                  55                  60
73 Glu Thr Cys Ala Lys Asn Cys Cys Leu Asp Gly Ala Ala Tyr Ala Ser
74 65                  70                  75                  80
75 Thr Tyr Gly Val Thr Thr Ser Gly Asn Ser Leu Ser Ile Gly Phe Val
76                 85                  90                  95
77 Thr Gln Ser Ala Gln Lys Asn Val Gly Ala Arg Leu Tyr Leu Met Ala
78             100                 105                 110
79 Ser Asp Thr Thr Tyr Gln Glu Phe Thr Leu Leu Gly Asn Glu Phe Ser
80         115                 120                 125
81 Phe Asp Val Asp Val Ser Gln Leu Pro Cys Gly Leu Asn Gly Ala Leu
82     130                 135                 140
83 Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro Thr
84 145                 150                 155                 160
85 Asn Thr Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln Cys
86                 165                 170                 175
87 Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly Trp
88             180                 185                 190
89 Glu Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Gly His Gly Ser
90         195                 200                 205
91 Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Ser Ile Ser Glu Ala
92     210                 215                 220
93 Leu Thr Pro His Pro Cys Thr Thr Val Gly Gln Glu Ile Cys Glu Gly
94 225                 230                 235                 240
95 Asp Gly Cys Gly Gly Thr Tyr Ser Asp Asn Arg Tyr Gly Gly Thr Cys
96                 245                 250                 255
97 Asp Pro Asp Gly Cys Asp Trp Asn Pro Tyr Arg Leu Gly Asn Thr Ser
98             260                 265                 270
99 Phe Tyr Gly Pro Gly Ser Ser Phe Thr Leu Asp Thr Thr Lys Lys Leu
100         275                 280                 285
101 Thr Val Val Thr Gln Phe Glu Thr Ser Gly Ala Ile Asn Arg Tyr Tyr
102     290                 295                 300
103 Val Gln Asn Gly Val Thr Phe Gln Gln Pro Asn Ala Glu Leu Gly Ser
104 305                 310                 315                 320
105 Tyr Ser Gly Asn Glu Leu Asn Asp Asp Tyr Cys Thr Ala Glu Glu Ala
106                 325                 330                 335
```

RAW SEQUENCE LISTING DATE: 11/18/2004
PATENT APPLICATION: US/10/804,785 TIME: 09:30:55

Input Set : A:\10804785.txt
Output Set: N:\CRF4\11182004\J804785.raw

```
107 Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln Phe
108                 340                 345                 350
109 Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp Asp
110         355                 360                 365
111 Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr Asn
112     370                 375                 380
113 Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr Ser
114 385                 390                 395                 400
115 Ser Gly Val Pro Ala Gln Val Glu Ser Gln Ser Pro Asn Ala Lys Val
116                 405                 410                 415
117 Thr Phe Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asn Pro
118             420                 425                 430
119 Ser Gly Gly Asn Pro Pro Gly Gly Asn Pro Pro Gly Thr Thr Thr Thr
120         435                 440                 445
121 Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln Ser
122     450                 455                 460
123 His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val Cys
124 465                 470                 475                 480
125 Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln Cys
126                 485                 490                 495
127 Leu
130 <210> SEQ ID NO: 3
131 <211> LENGTH: 1635
132 <212> TYPE: DNA
133 <213> ORGANISM: Hyprocrea orientalis
135 <400> SEQUENCE: 3
136 cgtcatctcg gccttcttgg ccacggcccg tgctcagtcg gcctgcactc tccaaacgga      60
137 gactcacccg tctctgacat ggcagaaatg ctcgtctggc ggcacttgca cccagcagac     120
138 aggctccgtg gtcatcgacg ccaactggcg ctggactcac gcgactaaca gcagcacgaa     180
139 ctgctacgac ggcaacactt ggagctcaac cctatgccct gacaacgaga cttgcgcgaa     240
140 gaattgctgc ctggacggtg ccgcctatgc gtccacgtac ggagtcacca cgagtgccga     300
141 cagcctctcc atcggcttcg tcacgcaatc tgcacagaag aacgttggcg cccgtctcta     360
142 cctgatggcg agtgacacga cttaccagga gttcacgctg cttggcaacg agttctcttt     420
143 tgacgttgat gtttcgcagc tgccgtaagt gacaaccatt ccccgcgagg ccatcttctc     480
144 attggttccg agctgacccg ccgatctaag atgtggcttg aacggcgctc tgtacttcgt     540
145 gtctatggat gcggatggtg gcgtgagcaa gtatcccacc aacaccgccg gcgccaagta     600
146 cggcacgggc tactgcgaca gccagtgccc ccgcgatctc aagttcatca acggccaggc     660
147 caacgttgaa ggctgggagc cgtcctccaa caacgccaac acgggtattg gcggacacgg     720
148 aagctgctgc tctgagatgg atatctggga ggccaactcc atctccgagg ctctgactcc     780
149 tcacccttgc acgactgttg gccaggagat ctgcgacggt gacggctgcg gcggaaccta     840
150 ctccaacgac cgatatggtg gtacttgcga tcctgatggt tgtgattgga atccataccg     900
151 cttgggcaac accagcttct atggccctgg ctcgagcttc accctcgata ccaccaagaa     960
152 gttgaccgtt gtcacccagt tcgagacctc gggtgccatc aaccgttact atgtccagaa    1020
153 cggcgtcact taccagcaac ccaacgccga gctcggtagt tactctggta atgagctcaa    1080
154 cgatgactac tgcacagctg aggagtcgga attcggcggc tcctccttct cggacaaggg    1140
155 cggccttact cagttcaaga aggccacttc cggcggcatg gtcctggtca tgagcttgtg    1200
156 ggatgacgtg agttgataga cagcattcac attgtcgttg gaaagacggg cggctaaccg    1260
157 agacatatga tatctaacag tactacgcca acatgctgtg gctggactcc acctacccga    1320
158 caaacgagac ctcctccacc cccggcgccg tgcgcggaag ctgctccacc agctccggcg    1380
```

RAW SEQUENCE LISTING DATE: 11/18/2004
PATENT APPLICATION: US/10/804,785 TIME: 09:30:55

Input Set : A:\10804785.txt
Output Set: N:\CRF4\11182004\J804785.raw

```
159 tccccgctca gctcgagtcc cagtccccca acgccaaggt cgtctactcc aacatcaagt    1440
160 tcgggcccat tggcagcacc ggcaacccca gcggcggaaa ccctcctggc ggaaaccctc    1500
161 ccggcaccac caccacccgc cgcccagcta ccaccactgg aagctctccc ggacctactc    1560
162 agactcacta cggccagtgc ggcggcatcg gctacagcgg ccctacggtc tgcgccagcg    1620
163 gcacgacctg ccagg                                                    1635
165 <210> SEQ ID NO: 4
166 <211> LENGTH: 17
167 <212> TYPE: PRT
168 <213> ORGANISM: Hyprocrea orientalis
170 <400> SEQUENCE: 4
171 Met Tyr Arg Lys Leu Ala Val Ile Ser Ala Phe Leu Ala Thr Ala Arg
172  1               5                  10                  15
173 Ala
176 <210> SEQ ID NO: 5
177 <211> LENGTH: 497
178 <212> TYPE: PRT
179 <213> ORGANISM: Hyprocrea orientalis
181 <400> SEQUENCE: 5
182 Gln Ser Ala Cys Thr Leu Gln Thr Glu Thr His Pro Ser Leu Thr Trp
183  1               5                  10                  15
184 Gln Lys Cys Ser Ser Gly Gly Thr Cys Thr Gln Gln Thr Gly Ser Val
185             20                  25                  30
186 Val Ile Asp Ala Asn Trp Arg Trp Thr His Ala Thr Asn Ser Ser Thr
187         35                  40                  45
188 Asn Cys Tyr Asp Gly Asn Thr Trp Ser Ser Thr Leu Cys Pro Asp Asn
189     50                  55                  60
190 Glu Thr Cys Ala Lys Asn Cys Cys Leu Asp Gly Ala Ala Tyr Ala Ser
191 65                  70                  75                  80
192 Thr Tyr Gly Val Thr Thr Ser Ala Asp Ser Leu Ser Ile Gly Phe Val
193                 85                  90                  95
194 Thr Gln Ser Ala Gln Lys Asn Val Gly Ala Arg Leu Tyr Leu Met Ala
195             100                 105                 110
196 Ser Asp Thr Thr Tyr Gln Glu Phe Thr Leu Leu Gly Asn Glu Phe Ser
197         115                 120                 125
198 Phe Asp Val Asp Val Ser Gln Leu Pro Cys Gly Leu Asn Gly Ala Leu
199     130                 135                 140
200 Tyr Phe Val Ser Met Asp Ala Asp Gly Gly Val Ser Lys Tyr Pro Thr
201 145                 150                 155                 160
202 Asn Thr Ala Gly Ala Lys Tyr Gly Thr Gly Tyr Cys Asp Ser Gln Cys
203                 165                 170                 175
204 Pro Arg Asp Leu Lys Phe Ile Asn Gly Gln Ala Asn Val Glu Gly Trp
205             180                 185                 190
206 Glu Pro Ser Ser Asn Asn Ala Asn Thr Gly Ile Gly Gly His Gly Ser
207         195                 200                 205
208 Cys Cys Ser Glu Met Asp Ile Trp Glu Ala Asn Ser Ile Ser Glu Ala
209     210                 215                 220
210 Leu Thr Pro His Pro Cys Thr Thr Val Gly Gln Glu Ile Cys Asp Gly
211 225                 230                 235                 240
212 Asp Gly Cys Gly Gly Thr Tyr Ser Asn Asp Arg Tyr Gly Gly Thr Cys
```

RAW SEQUENCE LISTING DATE: 11/18/2004
PATENT APPLICATION: US/10/804,785 TIME: 09:30:55

Input Set : A:\10804785.txt
Output Set: N:\CRF4\11182004\J804785.raw

```
213                 245                 250                 255
214 Asp Pro Asp Gly Cys Asp Trp Asn Pro Tyr Arg Leu Gly Asn Thr Ser
215             260                 265                 270
216 Phe Tyr Gly Pro Gly Ser Ser Phe Thr Leu Asp Thr Thr Lys Lys Leu
217         275                 280                 285
218 Thr Val Val Thr Gln Phe Glu Thr Ser Gly Ala Ile Asn Arg Tyr Tyr
219     290                 295                 300
220 Val Gln Asn Gly Val Thr Tyr Gln Gln Pro Asn Ala Glu Leu Gly Ser
221 305                 310                 315                 320
222 Tyr Ser Gly Asn Glu Leu Asn Asp Asp Tyr Cys Thr Ala Glu Glu Ser
223                 325                 330                 335
224 Glu Phe Gly Gly Ser Ser Phe Ser Asp Lys Gly Gly Leu Thr Gln Phe
225             340                 345                 350
226 Lys Lys Ala Thr Ser Gly Gly Met Val Leu Val Met Ser Leu Trp Asp
227         355                 360                 365
228 Asp Tyr Tyr Ala Asn Met Leu Trp Leu Asp Ser Thr Tyr Pro Thr Asn
229     370                 375                 380
230 Glu Thr Ser Ser Thr Pro Gly Ala Val Arg Gly Ser Cys Ser Thr Ser
231 385                 390                 395                 400
232 Ser Gly Val Pro Ala Gln Leu Glu Ser Gln Ser Pro Asn Ala Lys Val
233                 405                 410                 415
234 Val Tyr Ser Asn Ile Lys Phe Gly Pro Ile Gly Ser Thr Gly Asn Pro
235             420                 425                 430
236 Ser Gly Gly Asn Pro Pro Gly Gly Asn Pro Pro Gly Thr Thr Thr Thr
237         435                 440                 445
238 Arg Arg Pro Ala Thr Thr Thr Gly Ser Ser Pro Gly Pro Thr Gln Thr
239     450                 455                 460
240 His Tyr Gly Gln Cys Gly Gly Ile Gly Tyr Ser Gly Pro Thr Val Cys
241 465                 470                 475                 480
242 Ala Ser Gly Thr Thr Cys Gln Val Leu Asn Pro Tyr Tyr Ser Gln Cys
243                 485                 490                 495
244 Leu
247 <210> SEQ ID NO: 6
248 <211> LENGTH: 1589
249 <212> TYPE: DNA
250 <213> ORGANISM: Hyprocrea schweintzii
252 <400> SEQUENCE: 6
253 tcggcctgca ctctccaaac ggagactcac ccgtctctga catggcagaa atgctcgtct      60
254 ggcggcactt gcacccagca gacaggctcc gtggtcatcg acgccaactg gcgctggact     120
255 cacgctacta acagcagcac gaactgctac gacggcaaca cttggagctc aaccctgtgc     180
256 cctgacaatg agacttgcgc gaagaactgc tgcctggacg gtgccgccta tgcgtccacg     240
257 tacggagtca ccacgagtgc cgacagcctc tccatcggct tcgtgacaca gtctgcacag     300
258 aaaaacgttg gcgcccgtct ctacctgatg gcgagtgaca cgacttacca ggagttcacg     360
259 ctgcttggca acgagttctc attcgacgtt gatgtttcgc agctgccgta agtgacaacc     420
260 attcccccga cgccatcttc tcattggttc gaagctgacc cgccgatcta agatgtggct     480
261 tgaacggcgc tctttacttc gtgtccatgg acgcagatgg tggcgtgagc aagtatccca     540
262 ccaacaccgc cggcgccaag tacggcacgg gctactgtga cagccagtgc ccccgcgatc     600
263 tcaagtttat caacggccag gccaacgttg aaggctggga gccgtcctcc aacaacgcca     660
264 acacgggtat tggcggacac ggaagctgct gctccgagat ggatatctgg gaggccaact     720
```

RAW SEQUENCE LISTING ERROR SUMMARY DATE: 11/18/2004
PATENT APPLICATION: US/10/804,785 TIME: 09:30:56

Input Set : A:\10804785.txt
Output Set: N:\CRF4\11182004\J804785.raw

Please Note:
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:11; Xaa Pos. 273

VERIFICATION SUMMARY
PATENT APPLICATION: US/10/804,785

DATE: 11/18/2004
TIME: 09:30:56

Input Set : A:\10804785.txt
Output Set: N:\CRF4\11182004\J804785.raw

L:453 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:11 after pos.:272